

·综　述·

二代测序技术在微生物和感染性疾病中的应用

黄晶晶，肖　盟，徐英春

中国医学科学院　北京协和医学院　北京协和医院检验科，北京 100730

通信作者：徐英春　电话：010-69159766，E-mail：xycpumch@139.com

【摘要】随着感染性疾病防治在卫生保健方面的重要性日益凸显，感染性疾病引起的社会关注逐渐增多，新技术在人类预防和控制病原体传播方面发挥着巨大作用。微生物实验室作为病原体检测的一线科室，通过镜检、培养、鉴定、药敏等方法，在感染控制中发挥着重要作用。传统分子诊断和基因分型方法提供的信息有限，通常不能满足疫情爆发和传播调查需求。二代测序（next-generation sequencing，NGS）在单次序列测定中可确定菌株基因组完整的 DNA 序列，并从这些数据中得到抗菌药物耐药性、毒力及分型等可用于疫情调查的信息，进一步用于开展疫情特异性筛查。本文概述 NGS 技术及其在医院感染性疾病爆发调查、未知病原体鉴定、毒力分析、耐药基因组研究等临床微生物领域中的应用。

【关键词】二代测序；临床微生物学；感染性疾病
【中图分类号】R446.5　　【文献标志码】A

Application of Next Generation Sequencing in Clinical Microbiology and Infectious Diseases

HUANG Jing-jing, XIAO Meng, XU Ying-chun

Department of Clinical Laboratory, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China

Corresponding author: XU Ying-chun　Tel: 010-69159766, E-mail: xycpumch@139.com

【Abstract】 With the increasing importance of the prevention and treatment of infectious diseases in health care and the increasing social concern caused by infectious diseases, new technologies are crucial for preventing and controlling the spread of pathogens. The microbiology laboratory has always being playing a significant role as the first line of pathogen detection in infection control by carrying on smear microscopy, culture, identification, and susceptibility testing. Information provided by traditional molecular diagnostics and genotyping methods is limited, which can not satisfy the requirement of epidemiologic investigation on the outbreak and spread of communicable diseases. Next-generation sequencing determines the DNA sequence of a complete bacterial genome in a single sequencing run, from which information on resistance, virulence and typing is obtained. It is useful for investigation into the outbreak. The obtained genome data can be further used for developing an outbreak-specific screening test. In this review, a general introduction to next-generation sequencing and its applications in clinical microbiology including outbreak management, identification of unknown pathogens, taxonomy, and research on resistant genomes are presented.

【Key words】next-generation sequencing; clinical microbiology; infectious diseases

基金项目：中国医学科学院医学与健康科技创新工程项目（2016-12M-1-014）

随着新技术的发展，病原微生物鉴定方法已从传统的生化反应过渡到使用基质辅助激光解吸电离飞行时间质谱（matrix-assisted laser desorption/ionization time-of-flight mass spectrometry，MALDI-TOF MS）鉴定辅以双脱氧链终止法 Sanger 测序。Sanger 测序是菌种鉴定“金标准”，深受临床微生物工作者喜爱，但因其通量小，获得的序列信息有限，无法满足医院感染性疾病爆发调查、未知病原体鉴定、毒力分析及耐药性监测等方面的需求。21 世纪初人类基因组计划完成，测序技术发展迅猛，高通量、低成本的二代测序技术（next-generation sequencing，NGS）应运而生。与 Sanger 测序相比，NGS 不需要靶特异性引物，其优势在于单次运行即可用于所有病原体的鉴定和分型。因此，该技术可在医学微生物实验室和感染预防措施中发挥巨大作用[1]。

1 二代测序技术简介

NGS 即高通量测序技术，应用范围包括全基因组测序（whole-genome sequencing，WGS）、全外显子组测序（whole-exome sequencing，WES）、目标区域测序（targeted regions sequencing，TRS）和宏基因组测序。NGS 可在一次测试中同时运行多个病原体的 WGS，病原体可以是来自不同患者的病原分离物，也可以是来自某一个体的多个菌种临床样本（宏基因组学）。经过 10 多年的发展，NGS 测试成本明显降低[2-4]。该技术的核心思想是边合成边延伸边测序，通过捕捉合成链末端的信号（荧光信号或氢离子释放所致的 pH 值变化）来获得 DNA 序列信息。其主要流程是将通用接头连接到适当长度（一般 100~1000 bp）片段化的基因组 DNA 上，然后运用不同方法（油包水微球或固相表面连接等）形成千百万的单分子多克隆聚合酶链反应（polymerase chain reaction，PCR）阵列，继而进行大规模序列扩增并实时检测合成链末端的信号，再通过生物信息学分析得到完整的基因组序列信息[5]。现阶段发展成熟，应用较为广泛的技术平台主要包括 Roche/焦磷酸测序（也称 454 测序）、Illumina/Solexa 测序、Applied Biosystems/SOLID 测序和 Life Tech/Ion Torrent 测序等。

1.1 二代测序技术序列信息获取

NGS 序列的信息获取需要两个重要环节：（1）提取片段化的基因组 DNA。因测序仪可测的序列长度在 100~1000 bp 之间，故需在测序之前使用机械等方式将基因组 DNA 随机片段化处理，并在两端加上特定接头[5-6]；因各测序平台适用的片段 DNA 长度不一，且 DNA 片段大小对数据分析产生影响，所以应根据需要选择片段大小进行后续处理。（2）制备文库。DNA 或 RNA 片段需与接头和条形码融合，以便在测序后区分待测的 DNA 片段，随后进行克隆扩增，归一化和测序。为此，制备一个健全的包含正在被研究的基因组 DNA 或 RNA 的文库非常必要[7]。

1.2 二代测序技术数据分析

测序获得的大量序列信息需要通过生物信息学软件或在线平台进行分析，以得到完整的基因组序列。首先，可以使用诸如 FASTQC（Babraham Bioinformatics，英国）等程序总结序列质量信息，以确定数据质量合格。由于 NGS 具有高度敏感性，即使存在非目标生物体低水平污染也可能干扰下游分析，甚至导致不确定的分型结果。诸如 checkM 和 Kraken 等程序可用于快速筛选样品中可能存在的污染[8-9]。在某些情况下，生物信息学方法难以过滤污染物读数，样本则需重新分离和测序。

序列信息质量合格后，测序的片段（读数）可按需要进行组装（基因组装配）。一般来说，碎片越大，基因组装越容易和准确。常用的基因组组装软件有 CLC Genomics Workbench（Qiagen），Newbler 和 Velvet 等。同时，研究者可使用多位点序列分型（multilocus sequence typing，MLST），通过研究核心基因组多位点序列分型（core genome MLST，cgMLST）或全基因组多位点序列分型（whole genome MLST，wgMLST）研究分离株之间的遗传关系。这种方法可以使用诸如 SeqSphere（Ridom）和 BioNumerics（Applied Maths，Biomérieux）等软件或 EnteroBase 和 BIGSdb（细菌隔离基因组序列数据库）等在线工具[10]。

基因组装配后，基因组自动注释引擎可用于预测基因功能，从而提供生物体全面的功能图。通常使用的微生物基因组自动注释服务包括美国国立生物技术信息中心原核基因组注释传递途径（http://www.ncbi.nlm.nih.gov/genome/annotation_prok/）和 GeneMark（http://opal.biology.gatech.edu/）[11]，在基因组流行病学中心网站上可分析 NGS 数据。为了检测毒力和耐药基因，可使用 Virulence Finder 和 ResFinder，或通过综合抗菌药物耐药基因数据库和毒力因子数据库获得相应数据。然而，通过这些网站获得的结果需使用其他方法进行确认[12]。

使用 Artemis Comparison Tool（ACT）、Artemis 和

来自 Sanger Institute 的 DNA 绘图仪可进行进一步的比较基因组研究[13-15]。

2 二代测序技术在临床微生物领域的应用

如今，国内绝大部分临床微生物实验室采用基于生化反应的自动或半自动微生物鉴定系统进行菌种鉴定，虽然 MALDI-TOF MS 技术早在 20 世纪 80 年代末即已问世并迅速发展，但由于其设备昂贵尚未在众多基层医院广泛使用。中心实验室使用 MALDI-TOF MS 技术能够准确鉴定临床分离的细菌和真菌[16]，但对一些少见、罕见以及菌种复合体的鉴定结果并不十分可靠，如 Ranque 等[17]和 Wieser 等[18]报道的近平滑复合体、新型隐球菌复合体等。所以，临床微生物实验室常以 MALDI-TOF MS 技术辅以 Sanger 测序进行菌种鉴定。但 Sanger 测序获得的菌株信息有限，且临床对于微生物实验室的需求不仅仅局限于菌种鉴定和药物敏感性检测，实验室常常需要协助感染控制部门进行感染性疾病爆发的调查，在分子水平监测耐药性的传播。因此，及时、有效开展 NGS 测序是临床微生物工作必不可少的内容。

据报道，NGS 技术已应用于一些临床微生物实验室，其常用于医院感染性疾病爆发的调查、未知病原体的鉴定、毒力分析、耐药基因组的研究等方面[12,19]。

2.1 医院感染性疾病爆发的调查

NGS 序列信息可报告感染性疾病爆发的传播链，并已被用于跟踪由鲍曼不动杆菌、耐甲氧西林金黄色葡萄球菌和肺炎克雷伯杆菌引起的医院感染性疾病的传播[20-23]。在医院感染性疾病爆发期间，当流行病学证据支持时，基因组序列可用于跟踪患者之间或环境来源的传播链。当流行病学联系缺失时，基因组序列也可帮助作出医院感染来源的假设，并且确定哪些病例与医院感染有关。当 NGS 数据在医院感染调查中整合时，可指导、回应干预措施，例如隔离病房或制定新的防控指南，以防止未来爆发。

使用基于 WGS 分型的优点是可促进流行病学研究和公共卫生调查，其在爆发检测和监测多药耐药病原体的演变和动力学中显得尤为重要[12]。Wang 等[24]将 2013 年爆发于中国云南西双版纳的登革病毒进行 WGS 测序，经分析发现其属于 DENV-3 基因型Ⅱ，和老挝 2013 年分离株、孟加拉国 2002 年分离株基因型最为相似。Zhou 等[25]基于 WGS 描述了新出现的产 CTX-M-15 肺炎克雷伯菌序列型 1427（sequence type 1427，ST1427）。此外，通过基因组系统发育分析，追踪在单中心治疗患者和随后转诊患者之间间歇性传播的产 CTX-M-15 肺炎克雷伯菌 ST15。长期循环调查该区域患者群体以早期发现肺炎克雷伯菌高危克隆[1]。

除了疫情追踪和表征外，WGS 还有助于控制措施的实施以避免耐药病原体的扩散。荷兰曾发生机构间传播的碳青霉烯酶耐药肺炎克雷伯菌的爆发流行，卫生部门通过将所有检测阳性的居民隔离至医疗机构外单独治疗以避免更多人感染[26]。

2.2 未知病原体的鉴定

快速鉴定新发突发感染性疾病病原体，在最短时间内准确获取病原体的全面信息，对于能否有效控制感染性疾病疫情具有决定性意义。病毒性病原体的检测主要包括病毒抗原检测、核酸检测以及病毒分离培养等。病毒分离培养及抗原检测因方法学的固有缺陷，无法满足临床快速、准确鉴定新发突发病毒性病原体的需求[27-28]。虽然临床微生物实验室可采用特异性强、敏感性高的荧光定量 PCR 对疑似病毒进行筛查和确认，但其必须建立在已知病原体基因序列基础上，无法对未知病毒病原体进行检测[28]。

沙特阿拉伯 Saeb 等[27]在传统鉴定方法失败后，通过 NGS 技术结合生物信息学、系统发育和病原学分析报道了首例人感染溶血梭菌（*Clostridium haemolyticum*）。研究发现，该菌株是一种潜在的致病菌，具有抗生素抗性，且对人类宿主及其周围环境耐受力强，除质粒的新型重组外，染色体中存在毒力因子，可能是人类感染发生的原因。这项工作表明 NGS 用于临床微生物病原体鉴定，特别是对更多微生物样本进行 NGS 序列分析并在公共数据库中共享的重要性。

2.3 毒力分析

了解病原体毒力概况对于预测疾病严重程度和转归，以及在疾病早期进行风险评估至关重要。WGS 有可能通过使用在线工具来确定毒力因子的存在，不仅限于特定基因[29-30]。

WGS 可用于表征高毒力细菌，如产志贺毒素的大肠埃希菌（Shiga toxin-producing *Escherichia coli*，STEC）O104：H4 型。STEC 曾引发过大规模疫情，然而其已知的进化历史和基因组多样性信息较少。使用 WGS 进行爆发和非爆发菌株系统发育学分析，为其后续研究提供了进化背景，并揭示谱系特异性标记，指示了选择性压力和适应性选择[31]。此外，产志贺毒素的肠聚集性大肠埃希菌（enteroaggregative *Escherichia coli*，EAEC）Stx2a +O104：H4 的核心基因组系统发育分析显示，不

同聚类、不同耐药性、毒力模式取决于隔离时间[32]。

在大型队列研究中，WGS 可用于描述 STEC 的分子特征。所有关联菌株的基因型、血清型、MLST、毒力、抗菌药物耐药谱和系统遗传背景等信息均可从序列数据中获取，以获得全面、高分辨率的分子特征。NGS 可在相对较短时间内详细描述和比较多种菌株。因此，WGS 在改进病原体分子流行病学监测中的作用不可否认。

2.4 耐药基因组研究

除了鉴定病原体之外，还可使用宏基因组学方法来研究耐药基因组。肠道细菌是抗菌药物耐药基因(antibiotic resistance gene，ARG)的已知储存库，抗菌药物治疗对肠道细菌耐药基因组产生影响，可导致水平基因转移和耐药菌的选择。一项蒂宾根大学的研究调查了 2 例经环丙沙星治疗 6 d 的患者，通过宏基因组学分析显示肠内存在 ARG，并提出分析抗菌药物选择压力的新方法，可用于医院比较治疗方案及评估其对肠道细菌耐药基因组的影响；临床医生需格外关注抗菌药物的选择，优先使用对患者肠道菌群选择性压力较小的抗菌药物，可能减少耐药菌株的传播[33]。

NGS 数据库中已有报道，在生命垂危和住院患者体内发现了质粒黏菌素耐药基因 mcr-1。欧洲国家通过回顾性研究对 NGS 数据库中的 mcr-1 基因进行调查，从而引发丹麦、德国和荷兰的病例检测[34-36]。在荷兰，2000 多例当地患者肠杆菌科菌株在数小时内被筛选，以揭示 mcr-1 基因的存在[40]。因此，已经存在的 NGS 数据可用于筛选新的抗菌药物耐药基因。

3 总结和展望

NGS 技术的发展使得人类更深入地了解微生物的变迁和进化，更全面、准确地描述菌株基因型、毒力、抗菌药物耐药谱和系统遗传背景等信息，通过使用独特的标志物进行疫情追踪，为医院感染性疾病的爆发调查奠定理论基础。此外，NGS 在临床微生物实验室中可标准化病原体的分型方法，用于分子诊断、感染预防、病原体鉴定和监测以及新型抗菌药物耐药基因检测。然而，随着 NGS 在医学中的广泛应用，其技术流程和数据分析对临床微生物实验室提出了更高要求。研究者不仅需处理比传统基因分型更多的数据量，还要结合应用生物信息学和流行病学等内容。此外，从国内外研究来看，建立一个全球范围内共享的 NGS 数据库非常有必要，其将有助于突发、新发感染性疾病的及时诊断，并有助于开发微生物基因组学应用于感染性疾病监测和疫情调查的全部潜力。

然而，NGS 工作流程的改善，特别是缩短文库制备时间和数据分析自动化等方面尚需进一步探索。此外，建立统一的病原体分型标准，从而构建全球共享的 NGS 参考数据库以进一步降低成本势在必行。唯有如此，感染性病原体的监测与管理，以及基于 NGS 数据的靶向抗生素治疗才能实现，从而推动个性化微生物学的发展。

志谢：感谢中国医学科学院阜外医院周洲教授在 NGS 数据解读方面给予的指导；感谢北京协和医院检验科张栋在文章撰写方面给予的帮助。

参 考 文 献

[1] Zhou K, Lokate M, Deurenberg RH, et al. Use of whole-genome sequencing to trace, control and characterize the regional expansion of extended-spectrum beta-lactamase producing ST15 Klebsiella pneumoniae [J]. Sci Rep, 2016, 6: 20840.

[2] Dark MJ. Whole-genome sequencing in bacteriology: state of the art [J]. Infect Drug Resist, 2013, 6: 115-123.

[3] Sboner A, Mu XJ, Greenbaum D, et al. The real cost of sequencing: higher than you think! [J]. Genome Biol, 2011, 12: 125.

[4] Steuernagel B, Taudien S, Gundlach H, et al. De novo 454 sequencing of barcoded BAC pools for comprehensive gene survey and genome analysis in the complex genome of barley [J]. BMC Genomics, 2009, 10: 547.

[5] Junemann S, Sedlazeck FJ, Prior K, et al. Updating benchtop sequencing performance comparison [J]. Nat Biotechnol, 2013, 31: 294-296.

[6] Loman NJ, Misra RV, Dallman TJ, et al. Performance comparison of benchtop high-throughput sequencing platforms [J]. Nat Biotechnol, 2012, 30: 434-439.

[7] Head SR, Komori HK, LaMere SA, et al. Library construction for next-generation sequencing: overviews and challenges [J]. Biotechniques, 2014, 56: 61-64, 66, 68.

[8] Parks DH, Imelfort M, Skennerton CT, et al. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes [J]. Genome Res, 2015, 25: 1043-1055.

[9] Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments [J]. Genome Biol, 2014, 15: R46.

[10] Jolley KA, Maiden MC. BIGSdb: Scalable analysis of

bacterial genome variation at the population level [J]. BMC Bioinformatics, 2010, 11: 595.

[11] Aziz RK, Bartels D, Best AA, et al. The RAST Server: rapid annotations using subsystems technology [J]. BMC Genomics, 2008, 9: 75.

[12] Deurenberg RH, Bathoorn E, Chlebowicz MA, et al. Application of next generation sequencing in clinical microbiology and infection prevention [J]. J Biotechnol, 2017, 243: 16-24.

[13] Carver TJ, Rutherford KM, Berriman M, et al. ACT: the Artemis Comparison Tool [J]. Bioinformatics, 2005, 21: 3422-3423.

[14] Carver T, Harris SR, Berriman M, et al. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data [J]. Bioinformatics, 2012, 28: 464-469.

[15] Carver T, Thomson N, Bleasby A, et al. DNA Plotter: circular and linear interactive genome visualization [J]. Bioinformatics, 2009, 25: 119-120.

[16] 张小江，杨启文，王瑶，等. 2015 年北京协和医院细菌耐药性监测 [J]. 协和医学杂志，2016，7：334-341.

[17] Ranque S, Normand AC, Cassagne C, et al. MALDI-TOF mass spectrometry identification of filamentous fungi in the clinical laboratory [J]. Mycoses, 2014, 57: 135-140.

[18] Wieser A, Schneider L, Jung J, et al. MALDI-TOF MS in microbiological diagnostics-identification of microorganisms and beyond (mini review) [J]. Appl Microbiol Biotechnol, 2012, 93: 965-974.

[19] Tang P, Croxen MA, Hasan MR, et al. Infection control in the new age of genomic epidemiology [J]. Am J Infect Control, 2017, 45: 170-179.

[20] Halachev MR, Chan JZ, Constantinidou CI, et al. Genomic epidemiology of a protracted hospital outbreak caused by multidrug-resistant Acinetobacter baumannii in Birmingham, England [J]. Genome Med, 2014, 6: 70.

[21] Lewis T, Loman NJ, Bingle L, et al. High-throughput whole-genome sequencing to dissect the epidemiology of Acinetobacter baumannii isolates from a hospital outbreak [J]. J Hosp Infect, 2010, 75: 37-41.

[22] Harris SR, Cartwright EJ, Torok ME, et al. Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study [J]. Lancet Infect Dis, 2013, 13: 130-136.

[23] Koser CU, Holden MT, Ellington MJ, et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak [J]. N Engl J Med, 2012, 366: 2267-2275.

[24] Wang X, Ma D, Huang X, et al. Complete genome analysis of Dengue virus type 3 isolated from the 2013 Dengue outbreak in Yunnan, China [J]. Virus Res, 2017, 238: 164-170.

[25] Zhou K, Lokate M, Deurenberg RH, et al. Characterization of a CTX-M-15 producing *Klebsiella pneumoniae* outbreak strain assigned to a novel sequence type (1427) [J]. Front Microbiol, 2015, 6: 1250.

[26] Weterings V, Zhou K, Rossen JW, et al. An outbreak of colistin-resistant *Klebsiella pneumoniae* carbapenemase-producing *Klebsiella pneumoniae* in the Netherlands (July to December 2013), with inter-institutional spread [J]. Eur J Clin Microbiol Infect Dis, 2015, 34: 1647-1655.

[27] Saeb AT, Abouelhoda M, Selvaraju M, et al. The use of next-generation sequencing in the identification of a fastidious pathogen: a lesson from a clinical setup [J]. Evol Bioinform Online, 2017, 12: 1176934316686072.

[28] Guo C, Zhong LL, Yi HL, et al. [Clinical value of fluorescence lateral flow immunoassay in diagnosis of influenza A in children] [J]. Zhongguo Dang Dai Er Ke Za Zhi, 2016, 18: 1272-1276.

[29] Franz E, Delaquis P, Morabito S, et al. Exploiting the explosion of information associated with whole genome sequencing to tackle Shiga toxin-producing *Escherichia coli* (STEC) in global food production systems [J]. Int J Food Microbiol, 2014, 187: 57-72.

[30] Laabei M, Recker M, Rudkin JK, et al. Predicting the virulence of MRSA from its genome sequence [J]. Genome Res, 2014, 24: 839-849.

[31] Zhou K, Ferdous M, de Boer RF, et al. The mosaic genome structure and phylogeny of Shiga toxin-producing *Escherichia coli* O104: H4 is driven by short-term adaptation [J]. Clin Microbiol Infect, 2015, 21: 468.

[32] Ferdous M, Zhou K, de Boer RF, et al. Comprehensive characterization of *Escherichia coli* O104: H4 isolated from patients in the netherlands [J]. Front Microbiol, 2015, 6: 1348.

[33] Willmann M, El-Hadidi M, Huson DH, et al. Antibiotic selection pressure determination through sequence-based metagenomics [J]. Antimicrob Agents Chemother, 2015, 59: 7335-7345.

[34] Falgenhauer L, Waezsada SE, Yao Y, et al. Colistin resistance gene mcr-1 in extended-spectrum beta-lactamase-producing and carbapenemase-producing Gram-negative bacteria in Germany [J]. Lancet Infect Dis, 2016, 16: 282-283.

[35] Hasman H, Hammerum AM, Hansen F, et al. Detection of mcr-1 encoding plasmid-mediated colistin-resistant *Escherichia coli* isolates from human bloodstream infection and imported chicken meat, Denmark 2015 [J]. Euro Surveill, 2015, 20.

[36] Kluytmans-van den Bergh MF, Huizinga P, Bonten MJ, et al. Presence of mcr-1-positive Enterobacteriaceae in retail chicken meat but not in humans in the Netherlands since 2009 [J]. Euro Surveill, 2016, 21: 30149.

(收稿日期：2017-07-07)